Maths for ML

Nipun Batra

IIT Gandhinagar

August 15, 2025

Maths for ML

1. Given a vector of ϵ, we can calculate $\sum \epsilon_i^2$ using $\epsilon^T \epsilon$

Maths for ML

1. Given a vector of ϵ, we can calculate $\sum \epsilon_i^2$ using $\epsilon^T \epsilon$

$$\epsilon = \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \dots \\ \epsilon_N \end{bmatrix}_{N \times 1}$$

Maths for ML

1. Given a vector of ϵ, we can calculate $\sum \epsilon_i^2$ using $\epsilon^T\epsilon$

$$\epsilon = \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \dots \\ \epsilon_N \end{bmatrix}_{N\times 1}$$

$$\epsilon^T = \begin{bmatrix} \epsilon_1, \epsilon_2, \dots, \epsilon_N \end{bmatrix}_{1\times N}$$

Maths for ML

1. Given a vector of ϵ, we can calculate $\sum \epsilon_i^2$ using $\epsilon^T \epsilon$

$$\epsilon = \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \dots \\ \epsilon_N \end{bmatrix}_{N \times 1}$$

$$\epsilon^T = \begin{bmatrix} \epsilon_1, \epsilon_2, \dots, \epsilon_N \end{bmatrix}_{1 \times N}$$

$$\epsilon^T \epsilon = \sum \epsilon_i^2$$

Maths for ML

2.

$$(AB)^T = B^T A^T$$

Maths for ML

2.

$$(AB)^T = B^T A^T$$

3. For a scalar s

$$s = s^T$$

Maths for ML

4. Derivative of a scalar s wrt a vector θ

$$\theta = \begin{bmatrix} \theta_1 \\ \theta_2 \\ \vdots \\ \theta_N \end{bmatrix}$$

$$\frac{\partial s}{\partial \theta} =$$

Maths for ML

4. Derivative of a scalar s wrt a vector θ

$$\theta = \begin{bmatrix} \theta_1 \\ \theta_2 \\ \vdots \\ \theta_N \end{bmatrix}$$

$$\frac{\partial s}{\partial \theta} = \begin{bmatrix} \frac{\partial s}{\partial \theta_1} \\ \frac{\partial s}{\partial \theta_2} \\ \vdots \\ \frac{\partial s}{\partial \theta_N} \end{bmatrix}$$

Linear Functions: Row Vector Times Column Vector

Definition: Setup

Configuration:

- A is a row vector ($1 \times n$ matrix)
- θ is a column vector ($n \times 1$ matrix)
- $A\theta$ produces a scalar

Example: Concrete Example

$$\theta = \begin{bmatrix} \theta_1 \\ \theta_2 \end{bmatrix}_{2\times 1}, \quad A = \begin{bmatrix} A_1 & A_2 \end{bmatrix}_{1\times 2}$$

Key Points

Gradient of Linear Function: Key Result

Key Points

Computing the Gradient **Goal:** Find $\frac{\partial A\theta}{\partial \theta}$ where $A\theta = A_1\theta_1 + A_2\theta_2$

Example: Step-by-Step Calculation

$$\frac{\partial A\theta}{\partial \theta} = \begin{bmatrix} \frac{\partial}{\partial \theta_1}(A_1\theta_1 + A_2\theta_2) \\ \frac{\partial}{\partial \theta_2}(A_1\theta_1 + A_2\theta_2) \end{bmatrix}$$

$$= \begin{bmatrix} A_1 \\ A_2 \end{bmatrix}_{2\times 1} = A^T$$

Important: Fundamental Rule

Quadratic Forms and Their Derivatives

Quadratic Forms: Introduction

Definition: Quadratic Form Derivative Rule

Key Result: For matrix Z of form $X^T X$:

$$\frac{\partial}{\partial \theta}(\theta^T Z \theta) = 2Z^T \theta$$

Example: Understanding $X^T X$ Matrices

Starting with:

$$X = \begin{bmatrix} a & b \\ c & d \end{bmatrix}, \quad X^T = \begin{bmatrix} a & c \\ b & d \end{bmatrix}$$

Quadratic Forms: Introduction

Definition: Quadratic Form Derivative Rule

Key Result: For matrix Z of form X^TX:

$$\frac{\partial}{\partial\theta}(\theta^T Z\theta) = 2Z^T\theta$$

Example: Understanding X^TX Matrices

Starting with:

$$X = \begin{bmatrix} a & b \\ c & d \end{bmatrix}, \quad X^T = \begin{bmatrix} a & c \\ b & d \end{bmatrix}$$

Key Points

Computing $Z = X^TX$

Quadratic Forms: Introduction

Definition: Quadratic Form Derivative Rule

Key Result: For matrix Z of form X^TX:

$$\frac{\partial}{\partial\theta}(\theta^T Z\theta) = 2Z^T\theta$$

Example: Understanding X^TX Matrices

Starting with:

$$X = \begin{bmatrix} a & b \\ c & d \end{bmatrix}, \quad X^T = \begin{bmatrix} a & c \\ b & d \end{bmatrix}$$

Key Points

Computing $Z = X^TX$

Maths for ML

Let

$$Z = X^T X = \begin{bmatrix} e & f \\ f & g \end{bmatrix}_{2\times 2}$$

Maths for ML

Let

$$Z = X^T X = \begin{bmatrix} e & f \\ f & g \end{bmatrix}_{2\times 2}$$

$$\theta = \begin{bmatrix} \theta_1 \\ \theta_2 \end{bmatrix}_{2\times 1}$$

Maths for ML

Let

$$Z = X^T X = \begin{bmatrix} e & f \\ f & g \end{bmatrix}_{2\times 2}$$

$$\theta = \begin{bmatrix} \theta_1 \\ \theta_2 \end{bmatrix}_{2\times 1}$$

$$\theta^T Z \theta = \begin{bmatrix} \theta_1 & \theta_2 \end{bmatrix}_{1\times 2} \begin{bmatrix} e & f \\ f & g \end{bmatrix}_{2\times 2} \begin{bmatrix} \theta_1 \\ \theta_2 \end{bmatrix}_{2\times 1}$$

Maths for ML

Let

$$Z = X^T X = \begin{bmatrix} e & f \\ f & g \end{bmatrix}_{2\times 2}$$

$$\theta = \begin{bmatrix} \theta_1 \\ \theta_2 \end{bmatrix}_{2\times 1}$$

$$\theta^T Z \theta = \begin{bmatrix} \theta_1 & \theta_2 \end{bmatrix}_{1\times 2} \begin{bmatrix} e & f \\ f & g \end{bmatrix}_{2\times 2} \begin{bmatrix} \theta_1 \\ \theta_2 \end{bmatrix}_{2\times 1}$$

$$\theta^T Z \theta = \begin{bmatrix} \theta_1 & \theta_2 \end{bmatrix}_{1\times 2} \begin{bmatrix} e\theta_1 + f\theta_2 \\ f\theta_1 + g\theta_2 \end{bmatrix}_{2\times 1}$$

$$\theta^T Z \theta = e\theta_1^2 + 2f\theta_1\theta_2 + g\theta_2^2$$

The term $\theta^T Z \theta$ is a scalar.

Maths for ML

$$\begin{aligned}\frac{\partial}{\partial\theta}\theta^T Z\theta &= \frac{\partial}{\partial\theta}(e\theta_1^2 + 2f\theta_1\theta_2 + g\theta_2^2)\\ &= \begin{bmatrix}\dfrac{\partial}{\partial\theta_1}(e\theta_1^2 + 2f\theta_1\theta_2 + g\theta_2^2)\\ \dfrac{\partial}{\partial\theta_2}(e\theta_1^2 + 2f\theta_1\theta_2 + g\theta_2^2)\end{bmatrix}\\ &= \begin{bmatrix}2e\theta_1 + 2f\theta_2\\ 2f\theta_1 + 2g\theta_2\end{bmatrix} = 2\begin{bmatrix}e & f\\ f & g\end{bmatrix}\begin{bmatrix}\theta_1\\ \theta_1\end{bmatrix}\end{aligned}$$

$$= 2Z\theta = 2Z^T\theta$$

Matrix Rank and Invertibility

Matrix Rank: Fundamental Concept

Definition: What is Matrix Rank?

Rank = Maximum number of linearly independent rows (or columns)

Key Points

Two Equivalent Perspectives For an $r \times c$ matrix:

- **Row perspective:** r row vectors, each with c elements
- **Column perspective:** c column vectors, each with r elements

Example: Maximum Rank Rules

- If $r < c$: Maximum rank = r (more columns than rows)

Maths for ML: Matrix Rank

- Given a matrix A:

$$\begin{bmatrix} 0 & 1 & 2 \\ 1 & 2 & 1 \\ 2 & 7 & 8 \end{bmatrix}$$

Maths for ML: Matrix Rank

- Given a matrix A:

$$\begin{bmatrix} 0 & 1 & 2 \\ 1 & 2 & 1 \\ 2 & 7 & 8 \end{bmatrix}$$

- What is the rank?

Maths for ML: Matrix Rank

- Given a matrix A:

$$\begin{bmatrix} 0 & 1 & 2 \\ 1 & 2 & 1 \\ 2 & 7 & 8 \end{bmatrix}$$

- What is the rank?
- $r = c =$. Thus, rank is $<= 3$

Maths for ML: Matrix Rank

- Given a matrix A:

$$\begin{bmatrix} 0 & 1 & 2 \\ 1 & 2 & 1 \\ 2 & 7 & 8 \end{bmatrix}$$

- What is the rank?
- $r = c =$. Thus, rank is $<= 3$
- Row 3 can be written as: 3 times Row 1 + 2 times Row 1. Thus, Row 3 is linearly dependent on Row 1 and 2. Thus, rank(A)=2

Maths for ML: Matrix Rank

What is the rank of

$$X = \begin{bmatrix} 1 & 2 & 4 & 4 \\ 3 & 4 & 8 & 0 \end{bmatrix}$$

Maths for ML: Matrix Rank

What is the rank of

$$X = \begin{bmatrix} 1 & 2 & 4 & 4 \\ 3 & 4 & 8 & 0 \end{bmatrix}$$

Since X has fewer rows than columns, its maximum rank is equal to the maximum number of linearly independent rows. And because neither row is linearly dependent on the other row, the matrix has 2 linearly independent rows; so its rank is 2.

Maths for ML: Matrix Inverse

Suppose A is an $n \times n$ matrix. The inverse of A is another $n \times n$ matrix, denoted A^{-1}, that satisfies the following conditions.

$$\mathrm{A}\mathrm{A}^{-1} = \mathrm{A}^{-1}\mathrm{A} = \mathrm{I}_\mathrm{n}$$

Maths for ML: Matrix Inverse

Suppose A is an $n \times n$ matrix. The inverse of A is another $n \times n$ matrix, denoted A^{-1}, that satisfies the following conditions.

$$\mathrm{A}\mathrm{A}^{-1} = \mathrm{A}^{-1}\mathrm{A} = \mathrm{I_n}$$

where $\mathrm{I_n}$ is the identity matrix.

Maths for ML: Matrix Inverse

Suppose A is an $n \times n$ matrix. The inverse of A is another $n \times n$ matrix, denoted A^{-1}, that satisfies the following conditions.

$$\mathrm{A}\mathrm{A}^{-1} = \mathrm{A}^{-1}\mathrm{A} = \mathrm{I_n}$$

where $\mathrm{I_n}$ is the identity matrix.
Below, with an example, we illustrate the relationship between a matrix and its inverse.

Maths for ML: Matrix Inverse

Suppose A is an $n \times n$ matrix. The inverse of A is another $n \times n$ matrix, denoted A^{-1}, that satisfies the following conditions.

$$\mathrm{A}\mathrm{A}^{-1} = \mathrm{A}^{-1}\mathrm{A} = \mathrm{I_n}$$

where $\mathrm{I_n}$ is the identity matrix.
Below, with an example, we illustrate the relationship between a matrix and its inverse.

$$\begin{bmatrix} 2 & 1 \\ 3 & 4 \end{bmatrix} \begin{bmatrix} 0.8 & -0.2 \\ -0.6 & 0.4 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

$$\begin{bmatrix} 0.8 & -0.2 \\ -0.6 & 0.2 \end{bmatrix} \begin{bmatrix} 2 & 1 \\ 3 & 4 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

Maths for ML: Matrix Inverse

There are two ways to determine whether the inverse of a square matrix exists.

Maths for ML: Matrix Inverse

There are two ways to determine whether the inverse of a square matrix exists.

- If the rank of an $n \times n$ matrix is less than n, the matrix does not have an inverse.

Maths for ML: Matrix Inverse

There are two ways to determine whether the inverse of a square matrix exists.

- If the rank of an $n \times n$ matrix is less than n, the matrix does not have an inverse.
- When the determinant for a square matrix is equal to zero, the inverse for that matrix does not exist.

Maths for ML: Matrix Inverse

There are two ways to determine whether the inverse of a square matrix exists.

- If the rank of an $n \times n$ matrix is less than n, the matrix does not have an inverse.
- When the determinant for a square matrix is equal to zero, the inverse for that matrix does not exist.

Maths for ML: Matrix Inverse

There are two ways to determine whether the inverse of a square matrix exists.

- If the rank of an $n \times n$ matrix is less than n, the matrix does not have an inverse.
- When the determinant for a square matrix is equal to zero, the inverse for that matrix does not exist.

A square matrix that has an inverse is said to be nonsingular or invertible; a square matrix that does not have an inverse is said to be singular.

Maths for ML: Matrix Inverse

There are two ways to determine whether the inverse of a square matrix exists.

- If the rank of an $n \times n$ matrix is less than n, the matrix does not have an inverse.
- When the determinant for a square matrix is equal to zero, the inverse for that matrix does not exist.

A square matrix that has an inverse is said to be nonsingular or invertible; a square matrix that does not have an inverse is said to be singular.
Not every square matrix has an inverse; but if a matrix does have an inverse, it is unique.